

MGD+: An Enhanced Music Genre Dataset with Success-based Networks

Danilo B. Seufitelli, Gabriel P. Oliveira, Mariana O. Silva, Mirella M. Moro

Universidade Federal de Minas Gerais (UFMG) – Belo Horizonte, MG – Brazil

{daniloboechat, gabrielpoliveira, mariana.santos, mirella}@dcc.ufmg.br

Abstract. *Streaming platforms like Spotify have revolutionized music consumption, generating big volumes of data on hit songs. Such data serve as input to analyzing the music community and to the field of Music Information Retrieval. In this context, we present MGD+: an enhanced Music Genre Dataset with Success-based Networks. By combining Spotify chart data with acoustic metadata, we capture the evolution of musical careers. We further enhance the dataset with a genre-based collaboration network, represented as a graph, connecting artists through collaborations. MGD+ enables building success-based time series across several music markets, offers a friendly interface, and allows reproducibility; being a valuable tool for music-related tasks.*

1. Introduction

With the widespread availability of music streaming platforms and social media, large volumes of music data have become accessible for research and statistical purposes. These rich data sources offer unprecedented opportunities for the field of Music Information Retrieval (MIR) to delve into various aspects of music, such as analyzing musical characteristics, understanding listeners’ preferences, and studying trends and behaviors. Consequently, recently public available datasets allow to address relevant issues, mostly related to MIR [Silva et al. 2019b, Oliveira et al. 2021].

However, a common limitation of existing music datasets is their narrow scope, often centering on a single music market, such as the dominant US market. Whereas these datasets provide valuable insights into narrowed contexts, they may fail to capture the broader global music landscape and the significant influence of local and cultural aspects. Indeed, recognizing the importance of local artists, genres, and musical traditions is crucial for a comprehensive understanding of the music ecosystem [Oliveira et al. 2020, Oliveira and Moro 2023, Seufitelli et al. 2022]. Furthermore, the role of musical collaborations, which constantly transcend genre boundaries, is often overlooked in MIR research, despite being a prominent and influential factor [Silva et al. 2019a, Silva et al. 2023].

To address these limitations, we introduce MGD+ (*Music Genre Dataset*), an open dataset with temporal information on musical success in more than 60 countries, focusing on artist and genre careers. Specifically, we provide a chart-based success genre and artist network from 2017 to 2022. Such networks represent the collaboration between two perspectives: (the majority of) genres and artists. Besides, the dataset also contains metadata about the most relevant music elements, i.e., songs, acoustic metadata, artists, and genres. This novel set of features and its ease usability and reproducibility make

Table 1. Comparison of existing datasets with popularity data.

Year	Dataset	Reference	Size	Markets	Songs	Charts	Artists	Genres	Networks
2016	TPD	[Karydis et al. 2016]	23,385	1	✓	✓	✓	×	×
2019	HSPD	[Zangerle et al. 2019]	1,000,000	1	✓	✓	×	✓	×
2019	SPD	[Cosimato et al. 2019]	101,939	1	✓	×	✓	✓	×
2019	MusicOSet	[Silva et al. 2019b]	20,405	1	✓	✓	✓	✓	×
2020	4MuLA	[Silva et al. 2020]	96,458	1	✓	×	×	✓	×
2021	Unnamed	[Bertoni and Lemos 2021]	881	1	✓	×	×	×	×
2021	MUHSIC	[Oliveira et al. 2021]	22,635	1	✓	✓	✓	✓	×
2022	MUHSIC-BR	[Oliveira et al. 2022]	2,595	1	✓	✓	✓	✓	×
2020	MGD	[Oliveira et al. 2020]	13,880	9	✓	✓	✓	✓	✓
2023	MGD+	This work	126,213	68	✓	✓	✓	✓	✓

MGD+ a valuable resource for different MIR applications, such as Hit Song Science [Pachet 2011] and Music Genre Classification [Corrêa and Rodrigues 2016].

This paper is organized as follows. In Section 2, we provide an overview of the existing music-related datasets, highlighting their differences and how MGD+ addresses their shortcomings. Next, we present our new dataset in Section 3, through building methodology, its content, format and usage. In Section 4, we perform an exploratory analysis of the MGD+ content to gain insights into its characteristics. Then, we discuss possible scenarios and applications that illustrate the potential impact and usability of our dataset in MIR tasks in Section 5. Finally, in Section 6, we discuss the main challenges and limitations associated with the dataset, and in Section 7, we conclude by presenting ideas for further improvements and enhancements.

2. Existing Music Datasets

Over the years, the field of Music Information Retrieval has witnessed the development of music datasets aimed at simplifying research and analysis throughout distinct markets, genres and set of features. These datasets cover a wide range of attributes about music, from simple metadata to acoustic features, lyrics, and popularity. The specific subset of attributes included in each dataset varies depending on research and analysis goals. However, it is important to note that having access to a diverse and comprehensive collection of data enables a broader range of mining tasks and analytical opportunities. The availability of extensive music datasets empowers researchers and analysts to delve deeper into exploring and understanding of music-related phenomena.

Among many pieces of data, popularity is one of the most relevant aspects of the music industry. By incorporating popularity data into music datasets, researchers and analysts can study the patterns of success [Silva et al. 2023, Oliveira and Moro 2023], identify high-impact periods in artists’ careers [Barbosa et al. 2021, Seufitelli et al. 2022], and explore the dynamics of music consumption across different markets and timeframes. A better understanding of popularity enhances decision making and drives innovation within the music industry. Table 1 presents a selection of top datasets that incorporate popularity data alongside our MGD+ dataset, and their main features, including number of songs (*size*) and markets covered; data related to popularity charts, artists, genres; and networks built based on the dataset.

Specially, the Track Popularity Dataset (TPD) [Karydis et al. 2016] stands out as one of the pioneering datasets and contains track popularity data from 2004 to

2014, with tailored features for music information retrieval. The Hit Song Prediction Dataset (HSPD) [Zangerle et al. 2019] builds upon the Million Song Dataset (MSD) [Bertin-Mahieux et al. 2011] and includes data on tracks that made to the Billboard Hot 100 charts. Additionally, MusicOSet [Silva et al. 2019b] and SpotGenTrack Popularity Dataset (SPD) [Cosimato et al. 2019] focus on quality, with comprehensive metadata, lyrics, acoustic features, and song popularity data. Other datasets focus on local (regional) markets, such as Latin music in 4MuLA ([Silva et al. 2020]) and popular Brazilian songs in [Bertoni and Lemos 2021].

Among the all existing datasets, MUHSIC [Oliveira et al. 2021], MUHSIC-BR [Oliveira et al. 2022], and MGD [Oliveira et al. 2020] are the ones most similar to ours, as they have all been developed within our research group. MUHSIC and MUHSIC-BR focus on temporal success, providing comprehensive information on the temporal success of artists, songs, and genres – which aligns with the objective of our dataset. Similarly, the Music Genre Database (MGD) expands on the prospects of genre-related research and analysis by containing data on music genres, including genre collaboration networks. This overlaps with MGD+ dataset’s focus on artists and genre-based collaboration networks.

Still, the main difference and contribution of MGD+ over existing datasets lies both in the *scope* and the *coverage* of music data. Our dataset offers a wide range of attributes and gathers data from 68 different markets (international and regional). Its global coverage provides a comprehensive view of music consumption patterns and allows for comparative analyses across various types of culture.

3. MGD+ Dataset

In this section, we present MGD+ itself, an enhanced music genre dataset focused on success-based collaboration networks. First, we detail the steps of the dataset building methodology (Section 3.1). Then, we introduce the dataset content and the instructions to use it (Section 3.2).

3.1. Building Methodology

In recent years, there has been a profound shift in the way music is consumed worldwide, as people have transitioned from physical records to streaming services. This transformation has led to a significant impact on the global recorded music market, with streaming services emerging as the primary source of revenue since 2017. According to the International Federation of the Phonographic Industry (IFPI),¹ streaming services had a revenue of over 17.5 billion dollars in 2022, which represents approximately 67% of the total revenue generated by the music industry as a whole.

Therefore, in this dataset we consider data gathered from Spotify, the most popular audio streaming service worldwide. As of June 2023, it consists of more than 515 million users in over 180 markets.² Building upon previous research, considering several music markets (i.e., countries and territories) is crucial to understand the cultural differences between them and how each market behaves [Oliveira et al. 2020, Barbosa et al. 2021]. Thus, the data obtained from Spotify and enhanced in MGD+ helps to identify the unique factors influencing music consumption, as well as leads to a more nuanced understanding

¹State of the Industry Report 2023: <https://globalmusicreport.ifpi.org/>

²About Spotify: <https://newsroom.spotify.com/company-info/>

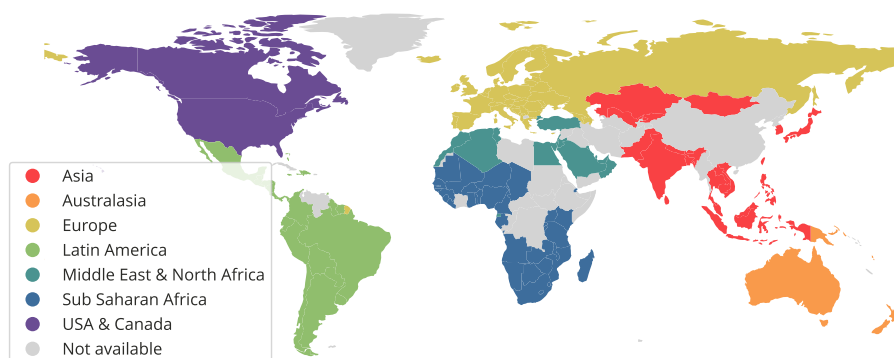


Figure 1. Spotify presence worldwide at the time of data collection (March 2022).

Table 2. Markets present in MGD+ per region.

Region	Markets
Asia	Hong Kong, India, Indonesia, Japan, Malaysia, Philippines, Singapore, South Korea, Taiwan, Thailand, Vietnam
Australasia	Australia, New Zealand
Europe	Austria, Belgium, Bulgaria, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Netherlands, Norway, Poland, Portugal, Romania, Russia, Slovakia, Spain, Sweden, Switzerland, Ukraine, United Kingdom
Latin America	Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru
Middle East & North Africa	Egypt, Israel, Morocco, Saudi Arabia, Turkey, United Arab Emirates
Sub-Saharan Africa	South Africa
USA & Canada	Canada, United States

of the global music landscape. Our building methodology then has three main steps: success charts collection, artist and song metadata acquisition, and network construction. All such steps are detailed next.

Daily Success Charts. Spotify provides daily and weekly success charts that include the Top 200 most streamed songs in each country in each day and week, respectively. To account for the diverse musical tastes exhibited by different countries, we have collected both global and regional daily charts spanning from January 2017 to March 2022. We did so by manually downloading the CSV files available on the Spotify Charts website. The data gathering has stopped in March 2022, because there have been significant changes to the Spotify Charts platform, and freely downloading the charts is no longer possible. In other words, a dataset such as MGD+ has become more important than ever, and shall be useful to any researcher, analyst, musician, producer, listener or company interested in such a part of music history.

As aforementioned, Spotify is available in more than 180 markets (see Figure 1). We have collected charts from 68 distinct markets worldwide, i.e., the markets that had CSV files available for download. The complete list of markets is listed in Table 2.

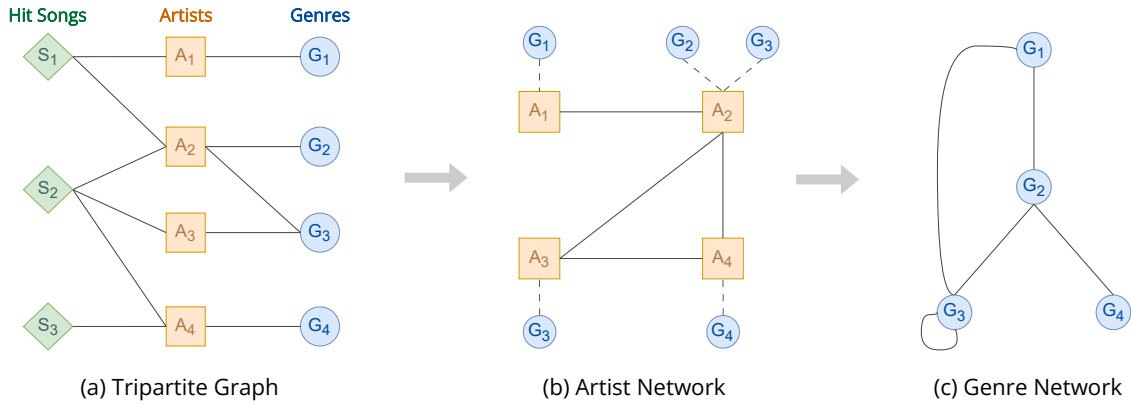


Figure 2. Reduction from the tripartite (a) to the one-mode Genre Collaboration Network (c). The intermediate step is an Artist Network with genre information (b). Artists and genres are linked when hit songs involve both nodes.

Hit Song and Artist Metadata. After collecting the hit songs (i.e., songs that are in the success charts) for each market, we use Spotify API³ to collect meaningful data about the hit songs and artists featured in the charts. This step means gathering data on all collaborating artists within a song, as the charts typically provide only the songs’ main artists. For the songs, we collect metadata such as their duration and a flag that indicates if the songs’ lyrics is explicit. Furthermore, we obtain acoustic features that describe the song itself, including the key and mode. Finally, for each artist, we obtain all the musical genres they are assigned to, as Spotify does not provide song or album genres.

Collaboration Networks. A Collaboration Network is typically represented as a graph consisting of nodes (vertices) that can be connected through edges. For instance, nodes represent artists, and an edge is formed between two artists if they have collaborated on a song. We model our collaboration networks following the methodology introduced in [Oliveira et al. 2020], which starts from the tripartite model composed of hit songs, artists, and genres (Figure 2a). In such a model, we equally consider all genres associated with an artist, as they contribute to shaping how fans and the music industry perceive the artist.

MGD+ provides two types of collaboration networks: artist and genre networks. The first one connects artists who collaborate (i.e., sing together) on one or more hit songs (Figure 2b). The genre information is not lost, as it is linked directly to the artists. Then, the genre networks connect the genres of artists who collaborate in the artist network. In such networks, the edges are undirected and weighted by the number of hit songs involving artists from both genres (Figure 2c). Also, it is possible to have self-loop edges that represent collaborations between artists from the same genre. In MGD+, we build both the artist and genre collaboration networks for each considered market and year.

3.2. Dataset Content and Usage

Overall, MGD+ contains daily charts from 68 distinct music markets, including 126,213 hit songs and 91,880 artists from 4,963 different music genres. To facilitate comprehension and enable further analysis, we structure MGD+ into a relational schema, depicted in Figure 3. This schema contains six main tables, with all the collected, curated, and en-

³Spotify API: <https://developer.spotify.com/>

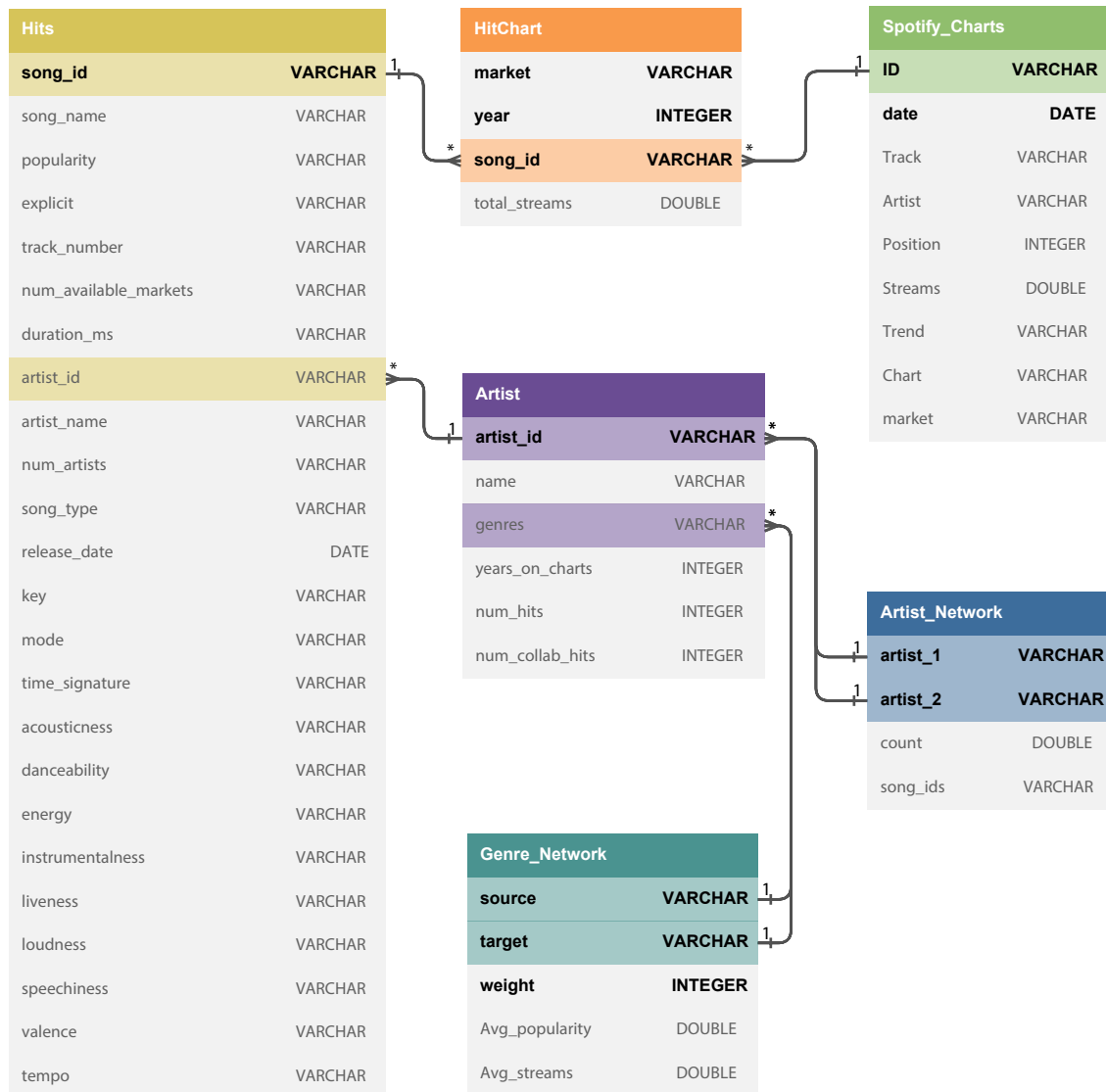


Figure 3. MGD+ relational schema (bold attributes stand for primary keys).

riched information. The tables represent hit songs, artists, charts (i.e., the main elements of the musical ecosystem), and the artist and genre networks.

MGD+ is publicly available in Zenodo, an open research data repository committed to the principles of open science [Seufitelli et al. 2023]. By making the dataset available, we encourage transparency, collaboration, and further data exploration by the research community. Finally, regarding its format, our dataset is available in **CSV files**, which simplifies data processing in popular programming languages such as Python and R, enabling the execution of complex analyses and visualizations.

4. Exploratory Data Analysis

In this section, we conduct an exploratory analysis of the MGD+ content to gain insights into its characteristics. We examine the charts from both global and 67 individual countries over a span of five years to analyze the diversity of genres present in each market. Figure 4 illustrates the number of distinct music genres extracted from artists who have

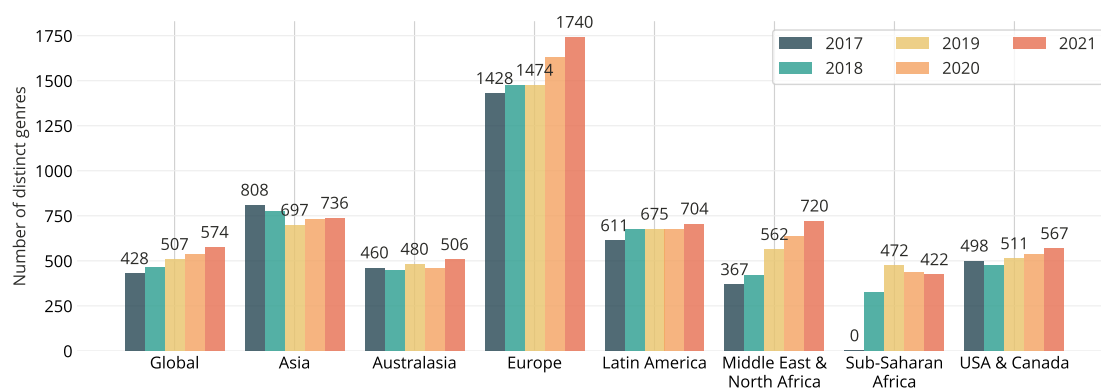


Figure 4. Number of distinct genres in Spotify charts for each region (2017-2021).

produced hit songs. The markets are grouped into seven cultural regions based on the IFPI classification. Across most markets (five out of seven), including the global scenario, there is an increasing number of distinct genres over time. This suggests that listeners in these markets are becoming more open to exploring and embracing new music styles.

Europe stands out as the region with the highest level of musical genre diversity. This diversity surpasses all other markets, including the global one. In 2021 alone, Europe recorded an impressive 1,740 registered genres. This finding underscores the richness and variety of musical expression within the European region. It means there is a thriving music scene by including various genres, catering to diverse listener preferences, and showcasing the cultural diversity present in Europe. In contrast, Sub-Saharan Africa and Australasia are the regions with the lowest levels of music genre diversity. This suggests a more concentrated music landscape, characterized by a smaller number of dominant genres or a narrower range of musical styles prevalent among the audience.

We also analyze music preferences by investigating the most popular genres in each market. Table 3 overviews the seven cultural regions and the global market. Notably, we observe that each country has its own distinct musical affinities; although, there is a prevalence of certain genres such as *pop*, *rap*, and *hip-hop* across many markets. The popularity of these genres can be justified, in part, due to the rising trend of collaboration songs among artists from different musical backgrounds. The fusion of genres and the cross-pollination of styles contribute to the appeal and widespread acceptance of genres like *pop*, *rap*, and *hip-hop* in diverse markets.

Nevertheless, it is important to note that while these genres dominate in terms of popularity, there is still room for a wide variety of musical preferences within each market. For instance, there are four regions with local genres on their top-5 list: Asia with *k-pop* and *j-pop*; Europe with *french hip hop*, *german hip hop*, and *dutch hip hop*; Latin America with *urbano latino*, *trap latino*, *reggaeton*, and *latin pop*; and Middle East & North Africa with *turkish hip hop* and *turkish pop*. These examples highlight the coexistence of global mainstream genres and the appreciation for local and regional music within each market. It emphasizes the diverse musical landscapes worldwide, with artists and listeners embracing and celebrating their cultural identities through music.

Table 3. Most popular music genres in each region from 2017 to 2022.

	Genre	Songs	Arts.		Genre	Songs	Arts.
<i>Global</i>	pop	2964	180	<i>Latin America</i>	urbano latino	4152	192
	rap	2535	137		trap latino	3603	145
	hip hop	1545	81		reggaeton	3429	148
	trap	1419	103		pop	2339	160
	melodic rap	898	40		latin pop	1835	137
<i>Asia</i>	pop	4464	205	<i>Middle East & North Africa</i>	pop	2433	179
	mandopop	2815	254		rap	1494	130
	k-pop	2436	200		turkish hip hop	1248	133
	j-pop	2090	181		turkish pop	1095	169
	cantopop	1377	164		hip hop	970	78
<i>Australasia</i>	pop	2609	187	<i>Sub-Saharan Africa</i>	rap	2164	138
	rap	1779	141		pop	1799	157
	hip hop	1147	89		hip hop	1306	82
	trap	785	102		trap	1127	104
	dance pop	549	96		melodic rap	706	37
<i>Europe</i>	french hip hop	5894	251	<i>USA & Canada</i>	rap	4772	153
	pop	5836	211		pop	3332	187
	pop urbaine	5577	187		trap	2943	133
	german hip hop	5244	284		hip hop	2708	98
	dutch hip hop	4004	240		melodic rap	1847	51

5. Applicability and Research Opportunities

In this section, we point out possible scenarios and applications that emphasize the potential impact and usability of MGD+ for several tasks in Music Information Retrieval, and related areas. Specifically, we discuss the following research opportunities: genre collaboration patterns in Section 5.1, exceptional genre patterns in Section 5.2, and collaboration recommendation in Section 5.3.

5.1. Genre Collaboration Patterns

As the landscape of the creative market continues to evolve, posing challenges for managing both predictive and diagnostic analyses becomes increasingly unpredictable. However, understanding the factors that contribute to an optimal musical partnership requires exploring collaboration patterns that directly influence its success [Baccigalupo et al. 2008, Bryan and Wang 2011, Silva et al. 2023]. Therefore, our new, rich dataset may help to uncover the dynamics of music cross-genre connections and music collaboration profiles within success-based networks. These networks consist of relationships formed by artists from different genres who come together to create hit songs. In fact, in [Oliveira et al. 2020], we show the importance of analyzing various music markets to ensure a comprehensive understanding of the music industry. We also recognize that focusing on data from markets other than the United States may represent the global music scenario more accurately. Therefore, adding even more markets (available in MGD+) promotes better knowledge about local genres’ influence as drive force in the music industry as well – even if locally.

5.2. Exceptional Genre Patterns

Combining artists from diverse genres (e.g., *Funk* with *Rock*) presents a relevant challenge when conducting descriptive analyses. At the same time, the increasing prevalence of collaborations exemplifies the inherently unpredictable nature of the music market. For example, in [Oliveira and Moro 2023], our objective is to delve deeper into exploring genre collaborations by using genre networks and collaboration profiles to uncover remarkable patterns in successful songs from recent years. We employ a methodology that combines social network analysis and data mining techniques to address the following research question:

In collaborative hit songs (i.e., featuring multiple artists), are there specific genre connection patterns that lead to above-average success?

Our new dataset is perfect for such a scenario, as our results show (see the aforementioned publication). Specifically, our findings unveil distinct exceptional subgroups within each market, highlighting the significance of considering local factors in success analyses. Hence, a fruitful analysis shall be expanding to a broader market to detect more exceptional local genre patterns.

5.3. Collaboration Recommendation

As another practical application, record labels may uncover frequent genre combinations that achieve a higher level of success to plan future song releases. Moreover, within the musical ecosystem, there is a compelling drive to explore previously uncharted and improbable patterns in genre mixtures. In such a challenging context, in [Oliveira 2021, Chapter 6] – which uses a previous simplified version of the dataset, *MGD*, we mine exceptional patterns of musical genres in songs that have been successful in both global and regional markets. In other words, we aim to verify if there is a relationship between combining different musical genres and success. Consequently, we discover and recommend collaborations between artists from different communities. Overall, *MGD+* may enrich the results by serving as input to identify newer patterns for all such markets.

6. Challenges and Limitations

In this section, we discuss the challenges encountered during the data collection process and highlight certain limitations that researchers should consider when using the *MGD+* for their analyses. While the dataset may provide valuable insights into music genres and success-based networks, it is important to acknowledge some challenges and limitations, discussed as follows.

Data Collection. The dataset collection presented several challenges. Due to the unavailability of direct access to Spotify’s charts, a crawler had to be implemented to gather the required data. This process required careful implementation and maintenance to ensure accurate and up-to-date data retrieval.

Collection Restrictions. Spotify no longer allows collecting charts data. As a result, the dataset may not capture the most recent trends and changes in music consumption patterns. Researchers should consider this limitation when analyzing the temporal evolution of musical careers or investigating *recent* market dynamics.

Regional Limitations. While the MGD+ provides data from 67 individual countries, it is essential to recognize that Spotify may not be the most popular music streaming service in all countries. Therefore, the dataset might not fully represent the music preferences and consumption patterns in regions where alternative platforms dominate. In particular, China, a significant music market according to the IFPI, is not available on Spotify. Therefore, researchers interested in analyzing the Chinese music market should seek alternative data sources to complement the MGD+.

Genre Representation. Another limitation to consider is that Spotify’s API provides information on the genres associated with artists rather than specific songs. While artist genres can provide insights into the overall musical style and preferences, analyzing song-specific genres could offer a more nuanced understanding of genre dynamics and trends.

Genre Label Accuracy. The genre labels provided by Spotify are sometimes derived using automated methods or subject to user-generated content. This can introduce inconsistencies and inaccuracies in the genre categorization, potentially impacting the reliability and precision of certain analyses that heavily rely on genre information.

7. Conclusion and Future Improvements

This paper introduced MGD+, an enhanced open dataset including metadata and musical success information regarding the main elements of the music ecosystem. By incorporating metadata and musical success information from charts spanning 67 countries over a five-year period, MGD+ offers a comprehensive view of the global music landscape. We modeled our dataset in a relational schema with six tables containing all the collected, curated, and enriched attributes from Spotify. Although MGD+ shares much content with other music-related datasets, its novelty relies on genre and artist networks, easily enabling a deeper understanding of the relationships and evolution of genres and artists. Such data is accessible and ready to use for complex tasks regarding Music Information Retrieval and Social Network Analyses, including music genre classification and hot streak prediction.

As future improvements, we first plan to expand the coverage of the dataset beyond Spotify by incorporating data from other popular music streaming platforms. This expansion shall allow to capture a more comprehensive view of the global music landscape and include countries where Spotify may not be the dominant platform. Second, we plan to enhance the genre information in the dataset by incorporating genre classification systems. Through considering alternative categorizations, we can provide a more diverse representation of musical genres, addressing the limitations of relying solely on Spotify’s genre classifications. By addressing the limitations, incorporating data from multiple platforms, and enhancing genre information, we aim to provide an even more comprehensive resource for music researchers, producers, engineers, analysts, among other music-related professionals.

Acknowledgments. This work was partially funded by CNPq, CAPES and FAPEMIG.

References

Baccigalupo, C., Plaza, E., and Donaldson, J. (2008). Uncovering affinity of artists to multiple genres from social behaviour data. In *ISMIR*, pages 275–280.

- Barbosa, G., Melo, B., Oliveira, G., Silva, M., Seufitelli, D., and Moro, M. (2021). Hot streaks in the brazilian music market: A comparison between physical and digital eras. In *SBCM*, pages 152–159. SBC. doi:10.5753/sbcm.2021.19440.
- Bertin-Mahieux, T. et al. (2011). The Million Song Dataset. In *Proc. of Int’l Society for Music Information Retrieval Conf. (ISMIR)*, pages 591–596.
- Bertoni, A. and Lemos, R. (2021). Três datasets criados a partir de um banco de canções populares brasileiras de sucesso e não-sucesso de 2014 a 2019. In *Anais do III Dataset Showcase Workshop*, pages 11–20. SBC. doi:10.5753/dsw.2021.17410.
- Bryan, N. J. and Wang, G. (2011). Musical influence network analysis and rank of sample-based music. In *ISMIR*, pages 329–334, Miami, USA.
- Corrêa, D. C. and Rodrigues, F. A. (2016). A survey on symbolic data-based music genre classification. *Expert Syst. Appl.*, 60:190–210. doi:10.1016/j.eswa.2016.04.008.
- Cosimato, A., Prisco, R. D., Guarino, A., Malandrino, D., Lettieri, N., Sorrentino, G., and Zaccagnino, R. (2019). The conundrum of success in music: Playing it or talking about it? *IEEE Access*, 7:123289–123298. doi:10.1109/ACCESS.2019.2937743.
- Karydis, I., Gkiokas, A., and Katsouros, V. (2016). Musical track popularity mining dataset. In *IFIP AIAI*, pages 562–572. doi:10.1007/978-3-319-44944-9_50.
- Oliveira, G. P. (2021). Analyses of musical success based on time, genre and collaboration. Master’s thesis, Universidade Federal de Minas Gerais, Brazil.
- Oliveira, G. P., Barbosa, G. R. G., Melo, B. C., Botelho, J. E., Silva, M. O., Seufitelli, D. B., and Moro, M. M. (2022). Musical Success in the United States and Brazil: Novel Datasets and Temporal Analyses. *Journal of Information and Data Management*, 13(1). doi:10.5753/jidm.2022.2350.
- Oliveira, G. P., Barbosa, G. R. G., Melo, B. C., Silva, M. O., Seufitelli, D. B., and Moro, M. M. (2021). MUHSIC: An Open Dataset with Temporal Music Success Information. In *SBB DSW*, pages 65–76, Rio de Janeiro, Brazil. doi:10.5753/dsw.2021.17415.
- Oliveira, G. P. and Moro, M. M. (2023). Exceptional collaboration patterns in music genre networks. In *BraSNAM*, pages 91–102. SBC. doi:10.5753/brasnam.2023.230516.
- Oliveira, G. P., Silva, M. O., Seufitelli, D. B., Lacerda, A., and Moro, M. M. (2020). Detecting collaboration profiles in success-based music genre networks. In *ISMIR*, Montreal, Canada.
- Pachet, F. (2011). Hit song science. In Tao Li, Mitsunori Ogihara, G. T., editor, *Music Data Mining*, chapter 10, pages 305–326. CRC Press, New York, NY, USA.
- Seufitelli, D. B., Oliveira, G. P., Silva, M. O., Barbosa, G. R., Melo, B. C., Botelho, J. E., de Melo-Gomes, L., and Moro, M. M. (2022). From Compact Discs to Streaming: A Comparison of Eras within the Brazilian Market. *Revista Vórtex*, 10(1). doi:10.33871/23179937.2022.10.1.2.
- Seufitelli, D. B., Oliveira, G. P., Silva, M. O., and Moro, M. M. (2023). MGD+: An Enhanced Music Genre Dataset with Success-based Networks. *Zenodo*. doi:10.5281/zenodo.8086642.

- Silva, A. C. M., Silva, D. F., and Marcacini, R. M. (2020). 4MuLA: A Multitask, Multimodal, and Multilingual Dataset of Music Lyrics and Audio Features. In *WebMedia*, page 145–148. ACM. doi:10.1145/3428658.3431089.
- Silva, M. O., Oliveira, G. P., Seufitelli, D. B., and Moro, M. M. (2023). Collaboration-aware hit song prediction. *Journal on Interactive Systems*, 14(1):201–214. doi:10.5753/jis.2023.3137.
- Silva, M. O., Rocha, L. M., and Moro, M. M. (2019a). Collaboration profiles and their impact on musical success. In *ACM SIGAPP*, page 2070–2077. ACM. doi:10.1145/3297280.3297483.
- Silva, M. O., Rocha, L. M., and Moro, M. M. (2019b). MusicOSet: An Enhanced Open Dataset for Music Data Mining. In *SBB DSW*, pages 408–417. SBC.
- Zangerle, E., Huber, R., and Yang, M. V. Y.-H. (2019). Hit Song Prediction: Leveraging Low- and High-Level Audio Features. In *ISMIR*, pages 319–326.